Unclassified
Security Classification

417841

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)
Northwestern University

2a. REPORT SECURITY CLASSIFICATION
Unclassified

2b. GROUP
Not applicable

3. REPORT TITLE
SOLUTION THEOREMS IN PROBABILISTIC PROGRAMMING: A LINEAR PROGRAMMING APPROACH

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
Technical Report

5. AUTHOR(S) (Last name, first name, initial)
Charnes, Abraham; Kirby, Michael J. L.; and Raike, William M.

6. REPORT DATE
March, 1967

7a. TOTAL NO. OF PAGES
26

7b. NO. OF REFS
24

8a. CONTRACT OR GRANT NO.
NONR 1228(10)

b. PROJECT NO.
NR 047-021 and:

c. Contract:
d. DA-31-124-ARO-D-322

9a. ORIGINATOR'S REPORT NUMBER(S)
Systems Research Memorandum No. 174

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
None

10. AVAILABILITY/LIMITATION NOTICES
Distribution of this document is unlimited.

11. SUPPLEMENTARY NOTES
None

12. SPONSORING MILITARY ACTIVITY
Logistics and Mathematical Statistics Branch
Office of Naval Research
U. S. Army Research Office

13. ABSTRACT

For some years research on solution theorems in probabilistic programming has been dormant. The obvious consequences of formal similarities to deterministic mathematical programming problems had been rapidly exhausted by researchers. Currently, however, the deeper study which was taking place during the "dormant period" has begun to produce results. On the one hand theorems characterizing optimal classes of stochastic decision rules for various general chance-constrained problems have been obtained. (See [11, 12, 13]). On the other hand, a great amount of effort has been expended on the special class of problems called linear programming problems under uncertainty (Cf. [2, 3, 4, 5, 6, 10, 16, 17, 18]), usually 2-stage and under still more special assumptions.

The general objective of these specializations has been to attain results and thereby to gain insight and technique to reapproach more fruitfully the more important and general but more recondite probabilistic programming problems. To this end, few abstractions or devices, from finite-dimensional Banach spaces ([3]) to the Kakutani fixed-point theorem ([5]) appear to have gone untried, except, perhaps, the ancient one of study and correlation of the existent results of other researchers.

It is the purpose of this paper to provide some such correlation and a redirection so that these simpler probabilistic programming problems may be overcome in <u>all</u> generality with new, simpler methods which offer some promise of extension to the more involved chance-constrained (and other probabilistic) models.

DD FORM 1 JAN 64 1473

Unclassified
Security Classification

Unclassified

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Probabilistic Programming						
Linear Programming Under Uncertainty						
Chance-Constrained Programming						
Constrained Medians						
Constrained Hypermedians						
Solvability and Existence Theorems						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through ________________."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through ________________."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through ________________."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, or *(U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, roles, and weights is optional.

SYSTEMS RESEARCH MEMORANDUM No. 174

The Technological Institute The College of Arts and Sciences

Northwestern University

SOLUTION THEOREMS IN PROBABILISTIC PROGRAMMING: A LINEAR PROGRAMMING APPROACH

by

A. Charnes, M. J. L. Kirby*, and W. M. Raike**

*Dalhousie University

** University of Texas and Northwestern University

March 1967

Part of the research underlying this report was undertaken for the Office of Naval Research, Contract NONR-1228(10), Project NR 047-021, and for the U. S. Army Research Office-Durham, Contract No. DA-31-124-ARO-D-322 at Northwestern University.

SYSTEMS RESEARCH GROUP

A. Charnes, Director

SOLUTION THEOREMS IN PROBABILISTIC PROGRAMMING:
A LINEAR PROGRAMMING APPROACH

by
A. Charnes,* M. J. L. Kirby,** and W. M. Raike***

Introduction

For some years research on solution theorems in probabilistic programming has been dormant. The obvious consequences of formal similarities to deterministic mathematical programming problems had been rapidly exhausted by researchers. Currently, however, the deeper study which was taking place during the "dormant period" has begun to produce results. On the one hand theorems characterizing optimal classes of stochastic decision rules for various general chance-constrained problems have been obtained. (See [11, 12, 13]). On the other hand, a great amount of effort has been expended on the special class of problems called linear programming problems under uncertainty (Cf. [2, 3, 4, 5, 6, 10, 16, 17, 18]), usually 2-stage and under still more special assumptions.

The general objective of these specializations has been to attain results and thereby to gain insight and technique to reapproach more fruitfully the more important and general but more recondite probabilistic programming problems. To this end, few abstractions or devices, from finite-dimensional Banach spaces ([3]) to the Kakutani fixed-point theorem ([5]) appear to have gone untried, except, perhaps, the ancient one of study and correlation of the existent results of other researchers.

It is the purpose of this paper to provide some such correlation and a redirection so that these simpler probabilistic programming problems may be overcome in <u>all</u> generality with new, simpler methods which offer some promise of extension to the more involved chance-constrained (and other

*Northwestern University
**Dalhousie University
*** University of Texas and Northwestern University

probabilistic) models. In so doing, we have been able to obtain significant extensions of past results as well as other new results which exhibit the power of this "Hessenberg" (here linear programming) as opposed to the previous "Kleinian" [1] approach.

Notation and Problem Statement

We begin with the 2-stage linear programming under uncertainty problem, which we write as follows: Determine, if they exist, a vector x_1 and a vector function $x_2(b_2)$ which minimize

$$(1) \qquad \begin{aligned} & c_1x_1 + E\,c_2x_2\,(b_2) \\ \text{subject to} \quad & A_{11}x_1 = b_1 \\ & A_{21}x_1 + A_{22}x_2(b_2) = b_2 \\ & x_1, x_2(b_2) \geq 0 \end{aligned}$$

In (1), c_1, c_2 and b_1 are vectors of known constants, A_{11}, A_{21} and A_{22} are known constant matrices, and b_2 is a vector of random variables with a known joint distribution function. The E operator denotes expectation, and the constraints involving $x_2\,(b_2)$ are required to hold with probability one. Juxtaposition of vectors or matrices indicates an inner product or matrix multiplication; whether vectors are to be row or column vectors

[1] "Klein" = elementary mathematics from a higher standpoint; "Hessenberg" = higher mathematics from an elementary standpoint.

will be clear from context. When needed, a superscript T will denote transposition, and for a vector y we will use $|y|$ to represent $(|y_1|, \ldots, |y_m|)$; a norm of y will be written $\|y\|$. It is a fundamental assumption of l. p. u. u., which we dub for convenience the "consistency" assumption, that for any $x_1 \geq 0$ satisfying $A_{11}x_1 = b_1$, there is an $x_2(b_2)$ satisfying the "second stage" constraints with probability one. To conform to past definitions (3) we specify the domain of definition of x_2 to be D_o, the convex hull of those subsets of Euclidean m-space which are assigned positive measure by the distribution of b_2.

To orient (1) in the framework of chance-constrained programming we note that the system $A_{11}x_1 = b_1$ may be considered to arise as the deterministic equivalent of chance constraints in which x_1 is to be chosen as a zero-order, or constant, decision rule.

Alternative Formulations

In [1] and later in [7], Charnes, Cooper and Thompson showed that (1) could be reduced to a deterministic "constrained hypermedian" problem involving x_1 alone, with "constrained medians" as the most important special type. The constrained hypermedian problem obtained there, which is equivalent to (1), is:

$$
(2) \qquad \begin{array}{ll} \text{minimize} & (c_1 - c_2 A_{22}^{\#} A_{21}) x_1 + E\left[\max\limits_{s=1,\ldots,K} w_s (A_{22}^{\#}(A_{21}x_1 - b_2)) \right] \\ \text{subject to} & A_{11}x_1 = b_1 \\ & x_1 \geq 0 \end{array}
$$

In (2), $A_{22}^{\#}$ is a generalized inverse of A_{22} [1] and the vectors w_s are

[1] For the definition of a von Neumann-Rao generalized inverse and a summary of some of its properties see [8] and [9].

the K extreme points of the convex polyhedral set defined by the constraints

$$wP = c_2 P$$
$$w \geq 0 \quad ,$$

where the matrix $P = I - A_{22}^{\#} A_{22}$ is the projection onto the null space of A_{22}. The computational (<u>not</u> theoretical) usefulness of (2) may be limited, since one might well require an enumeration of all the w_s. There are, however, a number of cases in which such an enumeration is not required.

A case of particular importance for economic theory (and also possibly for solution approximations in more general situations) is that in which, a priori, a linear programming problem for the first-stage gives an optimal solution to the 2-stage problem. We present a simple sufficient condition for this as the following theorem.

<u>Theorem 0:</u> If c_2 is in the range of the transpose of A_{22} (i.e., if $c_2 A_{22}^{\#} = c_2$, so that c_2 is a linear combination of the rows of A_{22}) then an optimal x_1 can be found by solving the linear programming problem

$$\text{(3)} \qquad \text{minimize } (c_1 - c_2A_{22}^{\#} A_{21})x_1 + \text{const. } (=Ec_2A_{22}^{\#} b_2)$$
$$\text{subject to} \qquad A_{11}x_1 = b_1$$
$$x_1 \geq 0$$

Proof: Since $c_2A_{22}^{\#} A_{22} = c_2$ we have $c_2x_2(b_2) = c_2A_{22}^{\#} A_{22}x_2(b_2)$, so the second stage constraints imply that for any feasible $x_2(b_2)$ (and we have assumed that one exists) $c_2x_2(b_2) = c_2A_{22}^{\#} (b_2 - A_{21}x_1)$. Taking expectations shows that the functional in (1) becomes that of (3), and the consistency assumption means that the second-stage constraints in (1) are redundant.

This is clearly one case in which no enumeration of extreme points is required, for (2) simply reduces to (3). Also, the theorem is particularly interesting when A_{22} is specialized to be square and nonsingular; it is then a simplification of previous results. (See [3], for instance.)

Another important such case is the "constrained median" type, which we render here as

$$\text{(4)} \qquad \text{minimize } c_1x_1 + \alpha\, E|b_2 - A_{21}x_1|$$
$$\text{subject to} \qquad A_{11}x_1 = b_1$$
$$x_1 \geq 0 \quad .$$

It can be assumed with no loss of generality that, in (4), $\alpha > 0$. We shall always assume, both here and in (1) and (2), that the system $A_{11}x_1 = b_1$ is consistent.

The terminology "constrained medians" is natural in the light of the absolute value in the functional of (4); for any random variable with a finite expectation, the median is the center about which the first absolute moment is minimized.

Wets, in [2] and [3], studied the special case of (1) in which $A_{22} = (I, -I)$, and (for unfathomable reasons) called this the "complete" problem. But he did not observe the connection with "constrained medians." We summarize in Theorem 1 below the relationship between the two problems; for reference we set forth the "complete problem" as:

$$
\text{(5)} \qquad \begin{array}{lll}
 & \text{minimize} & c_1x_1 + E\,[c_2y^+(b_2) + c_3y^-(b_2)] \\
\text{subject to} & & A_{11}x_1 = b_1 \\
 & & A_{21}x_1 + I\,y^+(b_2) - I\,y^-(b_2) = b_2 \qquad \text{w. p. 1} \\
 & & x_1 \geq 0 \\
 & & y^+(b_2), y^-(b_2) \geq 0 \qquad \text{w. p. 1}
\end{array}
$$

It is assumed in (5) that $c_2 + c_3 > 0$ and, of course, that $A_{11}x_1 = b_1$ is consistent.

<u>Theorem 1:</u> Any "complete problem" (5) can be represented as an equivalent "constrained median" problem (4), and conversely.

<u>Proof:</u> Set $\alpha = \frac{1}{2}(c_2 + c_3)$ and $c = c_1 - \frac{1}{2}(c_2 - c_3)A_{21}$. Then since $c_2y^+ + c_3y^- = \frac{1}{2}(c_2 + c_3)(y^+ + y^-) - \frac{1}{2}(c_2 - c_3)(y^- - y^+)$, using the constraints in (5) we obtain

$c_1x_1 + E\,[c_2y^+ + c_3y^-] = \alpha\, E\,[y^+ + y^-] + cx_1 + \frac{1}{2}(c_2 - c_3)\,Eb_2$.

Also, since $c_2 + c_3 > 0$ an optimal solution must clearly have $y^+y^- = 0$ so that $y^+(b_2) + y^-(b_2) = |b_2 - A_{21}x_1|$ at an optimum. Thus the functional in (5) becomes (dropping the constant term) $cx_1 + \alpha\, E|b_2 - A_{21}x_1|$, as asserted.

Conversely, taking $c_2 = c_3 = \alpha$ and $c_1 = c$, (4) leads immediately to (5). This completes the proof.

Having established Theorem 1, one may legitimately inquire as to how much more general than the constrained median problem the general l. p. u. u. problem (1) is. We shall show as a preliminary result in this direction that under very mild conditions on the distribution of b_2 the consistency assumption in (1) implies that the matrix A_{22} has a right inverse and therefore full row rank.

Theorem 2: Let (1) satisfy the consistency assumption. If the domain D_o contains $2m$ (where m is the number of entries in b_2) points $\{b_2^{(k)}, \hat{b}_2^{(k)}: k=1,\ldots,m\}$ such that, for each k, $\hat{b}^{(k)} = b^{(k)} + \Delta_k e_k$ for some real $\Delta_k \neq 0$ (in which e_k is the k-th unit vector), then A_{22} has a right inverse.

Proof: Let Δ be the diagonal matrix formed by the Δ_k values and let x_1 be any solution to $A_{11}x_1 = b_1$, $x_1 \geq 0$. Since there exists a feasible rule $x_2(b_2)$, we have $b^{(k)} - A_{22}x_2(b_2^{(k)}) = A_{11}x_1 = \hat{b}^{(k)} - A_{22}x_2(\hat{b}^{(k)})$ for all k. Letting H be the matrix whose k-th column is $x_2(b^{(k)}) - x_2(\hat{b}^{(k)})$, subtraction yields the relation

$$\Delta - A_{22} H = 0 .$$

But since $\Delta_k \neq 0$, Δ^{-1} exists; thus $A_{22}H\Delta^{-1} = I$, so that $H\Delta^{-1}$ is a right inverse for A_{22} which completes the proof.

It should be noted that Theorem 2 has an obvious generalization to the n-stage problem of linear programming under uncertainty.

Existence Theorems

In [5], Williams (with no acknowledgement of [1] or [7]) studied existence questions for the simple case $A_{11} = 0$, $b_1 = 0$, $A_{22} = (I, -I)$, with special assumptions on the distribution of b_2, by means of fixed-point methods; the results obtained there are special cases of those we shall give here. Further, we shall give, in our generality, an explicit characterization of what Williams termed the "insoluble finite" case, in which the objective function has a finite infimum but where there is no finite point at which the infimum is attained.

We proceed to establish necessary and sufficient conditions for the objective function in the constrained median problem (4) to have a finite infimum. For notational convenience we shall drop the subscript and use x in place of x_1, and will let $h(x) = cx + \alpha E|b_2 - A_{21}x|$.

We shall assume that b_2 has a finite expectation; the reason for this is expressed as Lemma 1.

Lemma 1: $h(x)$ is finite for each x if and only if $E|b_2|$ is finite.

Proof: Elementary inequalities on the absolute value yield the facts that

$$\pm|A_{21}x| \mp |b_2| \le \big||A_{21}x| - |b_2|\big| \le |b_2 - A_{21}x| \le |b_2| + |A_{21}x| .$$

Applying the expectation operator (which is order-preserving), we obtain

$$(6) \qquad (cx \pm \alpha|A_{21}x|) \mp \alpha E|b_2| \le h(x) \le (cx + \alpha|A_{21}x|) + \alpha E|b_2|$$

which establishes both sufficiency and necessity.

Theorem 3: There is a finite infimum $\bar{h}$ for $h(x)$ over the set $R = \{x: A_{11}x = b_1, \ x \ge 0\}$ if and only if there exist vectors w and v satisfying

$$(7) \qquad \begin{aligned} -wA_{21} + vA_{11} &\le c \\ |w| &\le \alpha . \end{aligned}$$

Proof: By virtue of inequality (6), it is enough to show that $cx + \alpha|A_{21}x|$ has a finite minimum for all x in the set R if and only if there exist vectors w and v satisfying the system (7). To do this, we note that the problem

$$\text{(8)} \qquad \begin{aligned} \text{minimize} \quad & cx + \alpha\,|A_{21}x| \\ \text{subject to} \quad & A_{11}x = b_1 \\ & x \geq 0 \end{aligned}$$

can be rewritten equivalently as the consistent linear programming problem

$$\text{(9)} \qquad \begin{aligned} \text{minimize} \quad & cx + \alpha\,(y^+ + y^-) \\ \text{subject to} \quad & A_{11}x = b_1 \\ & y^+ - y^- - A_{21}x = 0 \\ & x,\ y^+,\ y^- \geq 0 \ . \end{aligned}$$

The dual to (9) has as its objective function maximization of the functional vb_1; more importantly, the dual constraint set is exactly the system (7). Since we have assumed that the constraint set of (7) is consistent, the dual theorem of linear programming (see [14], vol. I, p. 190) immediately yields the desired result.

We now turn to an elucidation, in all generality, of what was termed in the special case of [5] the "insoluble finite" case. We assume the feasible region R is nonempty, $h(x)$ has a finite infimum over R, and yet there is no finite x which yields this infimum. Since $h(x)$ is clearly continuous, the only way this can happen is for there to exist a sequence $\{x^n\}$ of points in R, with $\|x^n\| \to \infty$ as $n \to \infty$, such that

$$\lim_{n\to\infty} h(x^n) = \inf_{x\in R} h(x) = \bar{h} \ .$$

It is clear that this case need never occur in practice since the set R can always be regularized by adjoining the constraint $\sum_j x_j \leq U$, where U is some large (possibly non-Archimedean) upper bound.

We will now show that there exist vectors $\hat{x}$, x_o such that

$$\lim_{t \to \infty} h(t\hat{x} + x_o) = \lim_{n \to \infty} h(x^n) = \inf_{x \in R} h(x) = \overline{h} ,$$

where $\|x^n\| \to \infty$ as $n \to \infty$ and $x^n \in R$ for all n. That is, there exists a displaced ray $t\hat{x} + x_o$ contained in R along which the limiting behavior of $h(t\hat{x} + x_o)$ as t become infinite imitates the limiting behavior of $h(x^n)$ as n becomes infinite.

We shall prove the following theorem via a sequence of lemmas; for notational simplicity we drop the subscripts and write A in place of A_{21} henceforth.

<u>Theorem 4:</u> Let $\{x^n : x^n \in R, \ n = 1,2,\dots\}$ be a sequence such that $\lim_{n \to \infty} h(x^n) = \inf_{x \in R} h(x) = \overline{h}$, and suppose that $h(x) > \overline{h}$ for all $x \in R$. Then there exist vectors $\hat{x}$ and x_o such that

$$\lim_{t \to \infty} h(t\hat{x} + x_o) = \lim_{n \to \infty} h(x^n) = \inf_{x \in R} h(x) = \overline{h} . \text{ Furthermore,}$$

$$c\hat{x} + \alpha |A\hat{x}| = 0 , \text{ and } x_o \in R \text{ and } t\hat{x} + x_o \in R \text{ for all } t \geq 0 .$$

<u>Lemma 2:</u> Let P be an $m \times k$ matrix and let P_o^n, P_o be m-vectors for $n = 1,2,\dots$ such that $P_o^n \to P_o$. If the systems $P\lambda^n = P_o^n$, $\lambda^n \geq 0$ all have solutions, then there is a solution λ to $P\lambda = P_o$ with $\lambda \geq 0$.

<u>Proof:</u> Suppose no such λ exists. Then by the Farkas lemma there is an m-vector u such that $uP_o < 0$ while $uP \geq 0$. But for all n,

$0 \leq (uP)\lambda^n = u(P\lambda^n) = uP_o^n$. This is impossible since $uP_o^n \to uP_o < 0$, proving the lemma.

Lemma 3: Let $J_o = \{j : x_j^n$ is bounded for all $n\}$. There is a subsequence of $\{x^n\}$ which we shall also denote by $\{x^n\}$ such that $x_j^n \to \bar{x}_j$ for some $\bar{x}_j \geq 0$ for all j in J_o and having $x_j^{n+1} > x_j^n$ for all j not in J_o. Further, the subsequence is such that the sign pattern of the $a^i x^n$ does not change with n.

Proof: The proof is trivial since the vectors x^n have only finitely many components. Henceforth we assume $\{x^n\}$ itself has the properties attributed to the subsequence; this causes no loss of generality.

Lemma 4: Define vectors $\tilde{x}^n$ and $\bar{x}$ as follows: Let $\tilde{x}^n$ have entries x_j^n for $j \notin J_o$ and zeros elsewhere; let $\bar{x}$ have entries $\bar{x}_j$ for $j \in J_o$ and zeros elsewhere. Let $s_i = \operatorname{sgn} a^i x^n$. Then

$$cx^n + \alpha|Ax^n| - c(\tilde{x}^n + \bar{x}) - \sum_i \alpha_i s_i a^i(\tilde{x}^n + \bar{x}) \to 0 .$$

Proof: $x^n - \tilde{x}^n - \bar{x} \to 0$ by construction, so that

$a^i x^n - a^i(\tilde{x}^n + \bar{x}) \to 0$ and $cx^n - c(\tilde{x}^n + \bar{x}) \to 0$. Also $\operatorname{sgn} a^i(\tilde{x}^n + \bar{x}) = s_i$ if $a^i x^n \not\to 0$, and if $a^i x^n \to 0$ then $a^i(\tilde{x}^n + \bar{x}) \to 0$ as well. The result follows.

Proof of Theorem: Without loss of generality, we can take as $\{x^n\}$ not only the subsequence obtained in Lemma 3, but may replace this by a further subsequence having the property that $cx^n + \alpha|Ax^n|$ converges to some limit which we denote by a. (This is true since $cx^n + \alpha|Ax^n|$ must stay bounded by virtue of inequality (3))..

In addition, we may assume that either $|a^i x^n| \to \infty$ or $a^i x^n$ converges to some finite limit, say $\bar{d}_i$; let $I_o = \{i : a^i x^n \to \bar{d}_i\}$.

Now for the sequence $\tilde{x}^n$ defined in Lemma 4,

$a^i\tilde{x}^n \to \pm\infty$ according as $a^i x^n$ does, and $a^i\tilde{x}^n \to \bar{d}_i - a^i\bar{x}$ when $a^i x^n \to \bar{d}_i$. Finally, $c\tilde{x}^n + \alpha|A\tilde{x}^n| \to a - \sum_i \alpha_i s_i a^i \bar{x}$, where a is as defined above.

Let us now apply Lemma 2 to the following system, where e_j denotes the j-th unit vector:

$$
\begin{aligned}
(10) \qquad & \sum_{j \notin J_o} a^i_j \lambda_j = \bar{d}_i - a^i\bar{x} , && i \in I_o \\
& \sum_{j \notin J_o} s_i a^i_j \lambda_j - \lambda_{n+i} = 1 , && i \notin I_o \\
& \sum_{j \notin J_o} (c_j + \sum_r \alpha_r s_r a^r_j)\lambda_j = a - \sum_r \alpha_r s_r a^r \bar{x} && \\
& \sum_{j \notin J_o} A_{11} e_j)\lambda_j = d - A_{11}\bar{x} && \\
& \lambda_{n+i}, \ \lambda_j \geq 0 . &&
\end{aligned}
$$

Lemma 2 guarantees from the behavior of the $\tilde{x}^n$ that there is a solution $\tilde{\lambda}$ to this system. Let us now define $\tilde{x}_j = \tilde{\lambda}_j$ for $j \notin J_o$ and $\tilde{x}_j = 0$ for $j \in J_o$ in order to obtain a vector $\tilde{x}$ having $\tilde{x}^n - \tilde{x} \geq 0$ for n sufficiently large; in fact $\tilde{x}^n - \tilde{x} \to \infty$ for $j \notin J_o$. The x_o promised in the statement of the theorem can now be defined as $x_o = \tilde{x} + \bar{x}$.

Now by construction of $\tilde{x}$ (via (7)), the vectors $\tilde{x}^n - \tilde{x}$ are all nonnegative and satisfy $a^i(\tilde{x}^n - \tilde{x}) \to 0$ for $i \in I_o$, $s_i a^i(\tilde{x}^n - \tilde{x}) \to \infty$ for $i \notin I_o$, and $c(\tilde{x}^n - \tilde{x}) + \sum_i \alpha_i s_i a^i(\tilde{x}^n - \tilde{x}) \to 0$. We can therefore apply Lemma 2 a second time to guarantee the existence of a solution $\hat{\lambda}$ to the system:

$$(11)\qquad \begin{aligned} \sum_{j \notin J_o} a^i_j \lambda_j &= 0, & i \in I_o \\ \sum_{j \notin J_o} s_i a^i_j \lambda_j - \lambda_{n+i} &= 1, & i \notin I_o \\ \sum_{j \notin J_o} (c_j + \Sigma \alpha_r s_r a^r_j) \lambda_j &= 0 \\ \sum_{j \notin J_o} A_{11} e_j) \lambda_j &= 0 \\ \lambda_j, \lambda_{n+i} &\geq 0 . \end{aligned}$$

Let us define the vector $\hat{x}$ by setting $\hat{x}_j = \hat{\lambda}_j$ for $j \notin J_o$ and $\hat{x}_j = 0$ for $j \in J_o$. The vector $\hat{x}$ thus defined has the properties that $a^i \hat{x} = 0$ for $i \in I_o$, $s_i a^i \hat{x} > 1$ for $i \notin I_o$, and that $c\hat{x} + \alpha|A\hat{x}| = c\hat{x} + \sum_i \alpha_i s_i a^i \hat{x} = 0$ as asserted by the theorem; also $A_{11}\hat{x} = 0$ so that $A_{11}(t\hat{x} + x_o) = A_{11}x_o = d$ for each $t \geq 0$.

We can also conclude from this construction that $a^i (t\hat{x} + x_o) = a^i x_o = \overline{d}_i$ for $i \in I_o$; $s_i a^i (t\hat{x} + x_o) = ts_i a^i \hat{x} + s_i a^i x_o$ which tends to infinity with t for $i \notin I_o$; and $c(t\hat{x} + x_o) + \alpha |A(t\hat{x} + x_o)| = t(c\hat{x} + \Sigma\alpha_i s_i a^i \hat{x}) + cx_o + \sum_i \alpha_i s_i a^i x_o = cx_o + \sum_i \alpha_i s_i a^i x_o = a$ for t sufficiently large to stabilize the sign pattern.

The proof is now finished as an easy consequence of the following theorem which serves to establish that $\lim_{t \to \infty} h(t\hat{x} + x_o) = \lim_{n \to \infty} h(x^n) = \overline{h}$.

<u>Theorem 5:</u> $\lim_{n \to \infty} h(x^n) = \lim_{n \to \infty} [cx^n + \sum_{i \notin I_o} \alpha_i s_i (a^i x^n - Eb_i)] +$

$$\sum_{i \in I_o} \alpha_i E|b_i - \overline{d}_i| .$$

Proof: Let $\chi_{in}(b_i) = \begin{cases} 1, & b_i < a^i x^n \\ 0, & b_i \geq a^i x^n \end{cases}$ for $i \notin I_o$ and all n.

Now $E|b_i - a^i x^n| = E[(b_i - a^i x^n)(1 - \chi_{in}(b_i))] +$

$$E[(a^i x^n - b_i)\,\chi_{in}(b_i)]$$

$$= [Eb_i - 2E(b_i \chi_{in}(b_i))] + a^i x^n [P(b_i < a^i x^n) - P(b_i \geq a^i x^n)] .$$

But $\lim_{n \to \infty} E(b_i \chi_{in}(b_i)) = \begin{cases} 0, & s_i = -1 \\ Eb_i, & s_i = +1 \end{cases}$, and the second term on the right gets as close as desired to $s_i a^i x^n$ for n sufficiently large. Thus $\lim_{n \to \infty} E|b_i - a^i x^n| = s_i(a^i x^n - Eb_i)$ for $i \notin I_o$; for $i \in I_o$ we have $\lim_{n \to \infty} a^i x^n = \bar{d}_i$, so that for such i, $\lim_{n \to \infty} E|b_i - a^i x^n| = E|b_i - \bar{d}_i|$. This proves theorem 5.

It should be emphasized that theorem 5 provides an explicit characterization of the limiting value of $h(x^n)$. We can also use theorem 5 to obtain the following result which further delimits those $\hat{x}$ which might be eligible in theorem 4.

Theorem 6: A necessary and sufficient condition that

$$\lim_{t \to \infty} h(tx + x_o) = \inf_{t \geq 0} h(tx + x_o) \text{ is that } cx + \alpha|Ax| = 0 .$$

Proof: Assume $\|x\| = 1$ since the conclusion is trivial for $x = 0$ and $cx + \alpha|Ax|$ is positive homogeneous. Theorem 5 and the fact that $E|b|$ is finite imply that $\lim_{t \to \infty} h(tx + x_o)$ exists and is finite (call the limit γ) if and only if $cx + \alpha|Ax| = 0$. Now necessity is obvious from theorem 5; to prove sufficiency, let $cx + \alpha|Ax| = 0$, and suppose there is a t_o such that $\inf_{t \geq 0} h(tx + x_o) = h(t_o x + x_o)$. Since $h(tx + x_o)$ is a convex function of t,

$$h(\tfrac{1}{2}[t_o x + x_o] + \tfrac{1}{2}[tx + x_o]) \leq \tfrac{1}{2}[h(t_o x + x_o) + h(tx + x_o)] \text{ for every t.}$$

But $\lim_{t \to \infty} h(tx + x_o) = \gamma = \lim_{t \to \infty} h(\tfrac{1}{2} t_o x + \tfrac{1}{2} tx + x_o)$, so that

$\gamma \leq \tfrac{1}{2} h(t_o x + x_o) + \tfrac{1}{2}\gamma$, or $\gamma \leq h(t_o x + x_o)$. Since $h(t_o x + x_o) \leq \gamma$ by definition of t_o, we have shown

$$\inf_{t \geq 0} h(tx + x_o) = h(t_o x + x_o) = \gamma = \lim_{t \to \infty} h(tx + x_o)$$ when t_o exists; if there is no such t_o then a fortiori $\inf_{t \geq 0} h(tx + x_o) = \lim_{t \to \infty} h(tx + x_o)$.

As prelude to a final result in this direction we note that, since $h(x)$ is convex, it has partial derivatives almost everywhere. It is of interest to establish asymptotic formulae for the radial directional derivative.

Lemma: $\lim_{t \to 0} \frac{d}{dt} h(tx) = cx + \alpha |Ax|$

Proof: Consider numbers $0 < t_1 < t_2 < t_3$. Let $\Delta_1 = \dfrac{h(t_2 x) - h(t_1 x)}{t_2 - t_1}$ and

$$\Delta_2 = \frac{h(t_3 x) - h(t_2 x)}{t_3 - t_2} \; .$$

Since h is convex, it follows easily if $\frac{d}{dt} h(tx)$ exists at $t = t_2$ that $\Delta_1 \leq \frac{d}{dt} h(t_2 x) \leq \Delta_2$. However, if we let $t_1, t_2, t_3 \to \infty$ in such a way that $t_2 - t_1$ and $t_3 - t_2$ stay constant, it is immediate from (13) below that Δ_1 and Δ_2 both approach the limit $cx + \alpha |Ax|$. The conclusion follows.

We wish to emphasize at this point that this "knife-edge" infinite displaced ray case can never appear as the solution to a practical problem. Rather it would exhibit an inadequacy or unrealistic formulation of the model. We have developed these results solely for the sake of completeness of analysis.

Probabilistic Interpretations and Incremental Formulae for h(x)

We now give expressions in terms of the probabilities of the various b versus Ax events for the increment of h(x) (i. e. $h(x+\xi) - h(x)$).

First, on inspecting figure 1 below, we note that, for $\epsilon > 0$,

$$f(u) = |u - (a+\epsilon)| - |u-a| = \begin{cases} \epsilon & u < a \\ \epsilon + 2(a-u), & a \le u < a+\epsilon \\ -\epsilon & u \ge a+\epsilon \end{cases} .$$

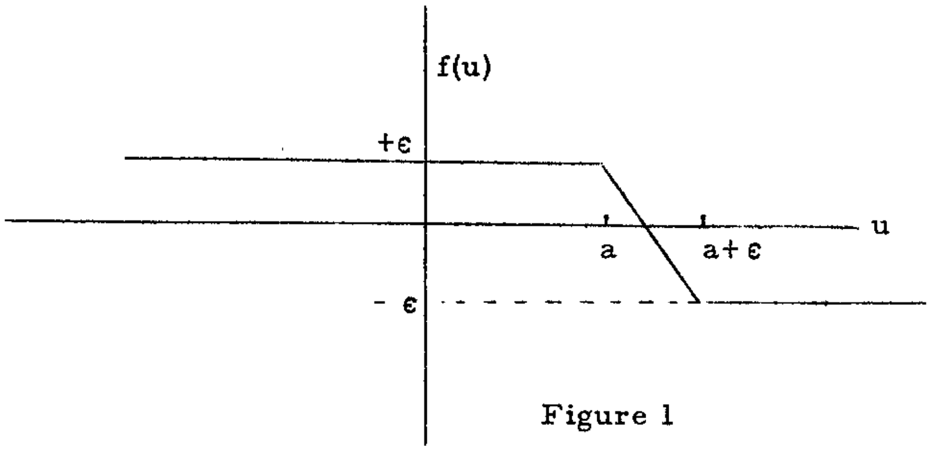

Figure 1

Thus, if u is a random variable with finite expectation,

$$E_u(|u-(a+\epsilon)| - |u-a|) = \epsilon\,[P(u<a) - P(u \ge a+\epsilon)] + [\epsilon + 2a - 2(a + \epsilon\,\theta(\epsilon))]\,P(a \le u < a+\epsilon)$$

for some $0 \le \theta(\epsilon) \le 1$ by the mean value theorem. In other words,

(12) $$E_u(|u-(a+\epsilon)| - |u-a|) = \epsilon\,[P(u < a+\epsilon) - P(u \ge a+\epsilon) - 2\theta(\epsilon)\,P(a \le u < a+\epsilon)].$$

But

$$h(x+\xi) - h(x) = c\,\xi + \alpha\,E(|b - Ax - A\xi| - |b - Ax|) ,$$

so if we use (12) we get, using $P(b \ge Ax)$ to denote the vector whose components are $P(b_i \geq a^i x)$, where a^i is the i-th row of the matrix A:

(13) $$h(x + \xi) - h(x) = c\,\xi + \alpha\, D(A\,\xi)\,[P(b < A(x + \xi)) - P(b \geq A(x + \xi)) - 2D(\theta(|A\,\xi|))P(Ax \leq b < A(x + \xi))] .$$

Here $D(y)$ denotes the diagonal matrix whose diagonal consists of the components of the vector y and $\theta(|A\,\xi|) = (\theta_1(|a^1\xi|), \ldots, \theta_m(|a^m\xi|))$ with $0 \leq \theta_i(|a^i\xi|) \leq 1$. Thus we have an expression for the increment of $h(x)$, and thereby an expression for any possible directional variation, in terms of the probabilities of various events.

Theorem 6: If the joint distribution function for b is continuous, the function $h(x)$ is continuously differentiable. The gradient of $h(x)$ is represented explicitly as:

(14) $$\nabla h(x) = c + w(x)\,A, \quad \text{where}$$
$$w_i(x) = \alpha_i\,[2F_i(a^i x) - 1]$$

Proof: In (13), let $\xi = t\,e_j$ where t is real and e_j is the j-th unit vector. The j-th component of $\nabla h(x)$ is then obtained as

$$\lim_{t \to 0} \frac{h(x + t\,e_j) - h(x)}{t} = c_j + \sum_i \alpha_i a_j^i\,[2F_i(a^i x) - 1].$$

This is the desired representation.

Corollary: If the distribution of b admits a continuous density function, or equivalently, if the distribution function is continuously differentiable, then the Hessian (the matrix of second partial derivatives) of $h(x)$ exists and is continuous. The jk-th entry is represented explicitly as

(15) $$2 \sum_i \alpha_i a_j^i a_k^i f_i (a^i x) \quad ,$$

where f_i is the marginal density of b_i .

It is possible to employ (13), (14) and (15) constructively to obtain efficient solution algorithms for the problem (4) and thereby for (5). We remark also that (14) is easily derived from the work in [2] and [5]. It is important to note, however, that (13) is valid for the general cases of discontinuous distributions. This result will be applied in further work elsewhere.

A major difficulty in applying existing nonlinear programming techniques to (4) directly is that multiple numerical integrations are required to calculate the value of $h(x)$ for any particular x. A notable exception is the projected gradient method of J. B. Rosen (see [20]) which requires only knowledge of the gradient at each step. Another possibility, suggested during a conversation by A. V. Fiacco in regard to the SUMT and related sequential unconstrained methods (see [21], [22], [23]), is that of performing either first-order or second-order gradient descent procedures (for a suitable unconstrained penalty function) only by reference to the gradient and inverted Hessian and not to the objective function value. Optimal step lengths for each (possibly mapped) gradient move can be calculated not only by explicitly minimizing the objective function along the desired vector, but rather by requiring that successive "moves" be orthogonal. Such a procedure appears to be novel, and will be reported on in more detail and with greater generality elsewhere.

Solvability Results for the General Problem.

The general N-stage linear programming under uncertainty problem (cf. [1], [6], [7], [11], [16]) can be written:

$$\text{(16)} \qquad \text{minimize } c_1x_1 + E\left(\sum_{j=2}^{N} c_jx_j\,(b_2,\ldots,b_j)\right)$$

$$\text{subject to} \qquad A_{i1}\,x_1 + \sum_{j=2}^{i} A_{ij}\,x_j\,(b_2,\ldots,b_j) = b_i, \qquad i=1,\ldots,N$$

$$x_1 \ge 0\ , \qquad x_j\,(b_2,\ldots,b_j) \ge 0\,, \qquad j=2,\ldots,N$$

In (16) the vectors b_1, $c_1,\ldots,c_n$ and the matrices A_{ij} are known constants; $b_2,\ldots,b_N$ are random vectors with a known joint distribution; and the decision rule x_j, for each j, is to be a function only of the random variables up to "stage" j, or equivalently is to be a constant function of b_{j+1} through b_N. The constraints which involve random variables are required to hold with probability one; we assume that the structure of the problem is such that, for any $x_1,\ldots,x_j$ satisfying the first j constraints in (16), there exists a decision rule $x_{j+1}(b_2,\ldots,b_{j+1})$ satisfying the $(j+1)$-st constraint (and the nonnegativity restrictions) with probability one. We shall refer to this as the "consistency assumption." We must specify the domain of definition of the functions x_j ; again following [3], we suppose that this domain D_o is the convex hull of those subsets of E_N which have positive probability under the distribution of the b_i.

We note again here that the consistency assumption implies that the matrices A_{ii} all have full row rank when the conditions of theorem 2 hold.

In [27] and [28], Wets established that there exists a convex programming problem which is a deterministic equivalent for the N-stage problem (in which the solution set for x_1 is a convex polyhedron); this result (obvious from (2) for $N = 2$) was extended by Murty in [10] where he showed that the nonzero components of an optimal x_1 correspond to linearly independent columns of the matrix $\begin{pmatrix} A_{11} \\ \vdots \\ A_{n1} \end{pmatrix}$.

Elaborate constructions, including the theory of polar cones, were used to obtain these results; we now show how they follow in a trivial manner simply by looking at (13) from a new point of view.

Theorem 7: If there exist any optimal decision rules $x_1^*, \ldots, x_N^*$ whatever for the N-stage problem (16), then there exists an optimal x_1 whose nonzero components correspond to linearly independent columns of the matrix $\begin{pmatrix} A_{11} \\ \vdots \\ A_{n1} \end{pmatrix}$.

Proof: Consider the optimal rules $x_j^* (b_2, \ldots, b_j)$ for $j = 2, \ldots, n$. Since there exists an x_1^* such that

$$A_{i1} x_1^* = b_i - \sum_{j=2}^{N} A_{ij} x_j^* (b_2, \ldots, b_j) \text{ for } i = 1, \ldots, N,$$ and since x_1^* is a constant, the right-hand expressions must also be constant functions of the random variables involved! Denote the vector of these constants by d^*, and consider the linear programming problem (which has a finite optimal value) given by:

$$\text{(17)} \qquad \begin{aligned} \text{minimize } & c_1 x_1 \\ A_{i1} x_1 &= d_i^* , \qquad i = 1, \ldots, N, \\ x_1 &\geq 0 . \end{aligned}$$

In (17), of course, $d_1^* = b_1$. Now there exists an optimal solution $\bar{x}_1$ to (17) which is a basic solution, i.e., whose nonzero components correspond to linearly independent columns of the constraint system.

Also, $(\bar{x}_1, x_2^*, \ldots, x_N^*)$ constitute an optimal solution to (16). To see this, note that the difference between the functional values in (16) given by this solution and the original solution $(x_1^*, \ldots, x_N^*)$ is $c_1\bar{x}_1 - c_1x_1^*$. Since the original solution is optimal for (16) we have $c_1\bar{x}_1 - c_1x_1^* \geq 0$, and since $\bar{x}_1$ is optimal for (17) we know that $c_1\bar{x}_1 \leq c_1x_1^*$; this completes the proof.

It may be possible to use the above observations as the basis for an efficient algorithm for (1), the reason being that in order to find d_2^* it is not necessary to know the entire function $x_2^*(b_2)$; $d_2^* = b_2 - A_{22}x_2^*(b_2)$ is known as soon as we know x_2^* for a single value of b_2. This approach will be particularly powerful if efficient methods can be found for determining $x_2^*(b_2)$ for some b_2, either explicitly or by means of approximations.

Appendix

We state here without proof (proofs may be found in [14]) three theorems which are of key importance in the theory of linear programming. Our lemma 2 is a restatement of part of the Farkas-Minkowski theorem in its homogeneous form; the key fact used in our proof of theorem 5 was the LIEP theorem. The Opposite Sign theorem is the heart of the proof of the Farkas-Minkowski theorem, and provides the theoretical basis for computational procedures for constructing basic solutions from nonbasic ones while not worsening the objective function value ; see [24].

It should be noted that all theorems may be proved without recourse to topological properties of the real numbers and are thus valid for vector spaces over arbitrary ordered fields.

A. Farkas-Minkowski Theorem: The intersection of a finite number of (possibly displaced) half-spaces, when a bounded set, is the convex hull of a finite number of extreme points.

Let $L = \{\lambda : P\lambda = P_o, \lambda \geq 0\}$; here and in the following let λ and α be n-vectors, P be m-by-n, and P_o be an m-vector. L is thus a polyhedral convex set.

B. LIEP Theorem (Linear Independence by Association with Extreme Points): $\lambda \neq 0$ is an extreme point of L if and only if the nonzero coordinates (components) of λ correspond to linearly independent columns of P.

C. Opposite Sign Theorem: L is spanned by its extreme points if and only if for every $\alpha \neq 0$ such that $P\alpha = 0$ it is necessary that one component of α be positive and that another component be negative.

We remark here that the Opposite Sign Theorem is also true in certain infinite-dimensional cases (specifically, in "generalized finite

[12] Charnes, A., and M. J. L. Kirby, "Optimal Decision Rules for the Triangular Chance-Constrained Programming Problem," Systems Research Memorandum No. 115, (Evanston, Ill., Northwestern University, 1965), Management Science (forthcoming).

[13] Charnes, A., and M. J. L. Kirby, "Some Special P-Models in Chance-Constrained Programming," Systems Research Memorandum No. 135, (Evanston, Ill., Northwestern University, August, 1965), Management Science (forthcoming).

[14] Charnes, A., and W. W. Cooper, Management Models and Industrial Applications of Linear Programming, 2 vols. (New York: John Wiley & Sons, Inc., 1961).

[15] Charnes, A., M. J. L. Kirby, and W. M. Raike, "On a Special Class of Constrained Generalized Median Problems," Systems Research Memorandum No. 155, (Evanston, Ill., Northwestern University, April, 1966).

[16] Van Slyke, R. M., and R. Wets, "Programming Under Uncertainty and Stochastic Optimal Control," (Seattle, Wash., Boeing Scientific Research Laboratories, Document Dl-82-0429, April, 1965).

[17] Van Slyke, R. M., and R. Wets, "L-Shaped Linear Programs with Applications to Optimal Control and Stochastic Programming," (Seattle, Wash., Boeing Scientific Research Laboratories, Document Dl-82-0526, April, 1966).

[18] Walkup, D. W., and R. Wets, "Stochastic Programs with Recourse," Seattle, Wash., Boeing Scientific Research Laboratories, Document Dl-82-0551, July, 1966.

[19] Charnes, A., and W. W. Cooper, "Deterministic Equivalents for Optimizing and Satisficing Under Chance Constraints," Operations Research 11, No. 1 (Jan.-Feb., 1963) pp. 18-39.

[20] Rosen, J. B., "The Gradient Projection Method for Nonlinear Programming, Part I; Linear Constraints," SIAM Journal 8, No. 1 (Mar. 1960), pp. 181-217.

[21] Fiacco, A. V. and G. P. McCormick, "The Sequential Unconstrained Minimization Technique for Nonlinear Programming, A Primal-Dual Method," Management Science 10, No. 2 (Jan. 1964) pp. 360-366.

[22] Fiacco, A. V. and G. P. McCormick, "Computational Algorithm for the Sequential Unconstrained Minimization Technique for Nonlinear Programming," Management Science 10, No. 4, (July, 1964) pp. 601-617.

[23] Fiacco, A. V., "Sequential Unconstrained Minimization Methods for Nonlinear Programming," Ph. D. Dissertation, Northwestern University, Evanston, Ill., June 1967.

[24] Charnes, A., K. O. Kortanek, and W. M. Raike, "Extreme-Point Solutions in Mathematical Programming: An Opposite-Sign Algorithm," Systems Research Memorandum No. 129 (June 1965).